DOI 10.1002/pmic.200401042 *Proteomics* 2005, *5,* 1624–1633

REGULAR ARTICLE

The wheat (*Triticum aestivum* L.) leaf proteome

Bryna E. Donnelly[1], Robin D. Madden[1], Patricia Ayoubi[2], David R. Porter[3] and Jack W. Dillwith[1]

[1] Department of Entomology & Plant Pathology, Oklahoma State University, Stillwater, OK, USA
[2] Department of Biochemistry & Molecular Biology, Oklahoma State University, Stillwater, OK, USA
[3] United States Department of Agriculture–Agricultural Research Service, Stillwater, OK, USA

The wheat leaf proteome was mapped and partially characterized to function as a comparative template for future wheat research. In total, 404 proteins were visualized, and 277 of these were selected for analysis based on reproducibility and relative quantity. Using a combination of protein and expressed sequence tag database searching, 142 proteins were putatively identified with an identification success rate of 51%. The identified proteins were grouped according to their functional annotations with the majority (40%) being involved in energy production, primary, or secondary metabolism. Only 8% of the protein identifications lacked ascertainable functional annotation. The 51% ratio of successful identification and the 8% unclear functional annotation rate are major improvements over most previous plant proteomic studies. This clearly indicates the advancement of the plant protein and nucleic acid sequence and annotation data available in the databases, and shows the enhanced feasibility of future wheat leaf proteome research.

Submitted: June 10, 2004
Revised: August 13, 2004
Accepted: September 20, 2004

Keywords:
Matrix-assisted laser desorption/ionization-time of flight / Plant proteomics / Protein database / Two-dimensional gel electrophoresis

1 Introduction

Wheat is an extremely important agronomic crop worldwide, with consumption doubling in the last 30 years to nearly 600 million tons per year (http://www.cimmyt.cgiar.org/Research/Economics/map/facts_trends/). The International Maize and Wheat Improvement Center has stated that the worldwide demand will increase over 40% by 2020, while the land and resources available for production will decrease significantly if current trends prevail [1]. Increased knowledge of wheat's biochemical constitution and functional biology is required to improve wheat in ways that will meet this demand. The Environmental Protection Agency, *via* mandates of the 1996 Food Quality Protection Act, is re-evaluating pesticides currently in use, and banning those with higher perceived risks to the environment and human health. The need for enhanced natural tolerance/resistance to biotic and abiotic stresses has never been greater, and will most likely be found through evaluation and elucidation of biochemical mechanisms already present in certain plant species and varieties. To this end, proteomic approaches can be utilized to ascertain target enzymes and proteins from resistant lines that could be utilized to enhance the natural tolerance of agronomically favorable varieties of plants. With this ultimate goal in mind, it is first necessary to develop approaches for the large-scale identification of wheat proteins.

Correspondence: Dr. Bryna E. Donnelly, Department of Entomology, The Pennsylvania State University, 505 Agricultural Sciences Building, University Park, PA 16802, USA
E-mail: bed11@psu.edu
Fax: +1-814-865-3048

Abbreviations: MADS, MCM1, agamous deficiens and serum response factor; **ME**, β-mercaptoethanol; **NBS-LRR**, nucleotide-binding site-leucine-rich repeat; **Rubisco**, ribulose bisphosphate carboxylase/oxygenase

Proteomic analysis of wheat endosperm proteins has been conducted to evaluate the end-product quality of wheat. Wheat seed storage proteins are partially responsible for dough quality, and hence have been evaluated extensively in an attempt to elucidate biochemical properties, which could be enhanced to yield higher quality dough [2]. Starch synthesis and accumulation, which aid in dough quality, occur during the grain-filling process. The timing, duration, and rate of grain-filling impact protein yield and thus dough quality. Factors affecting grain-filling are controlled by the amyloplasts, specialized leucoplasts found in the endosperm. Proteomic analysis of wheat amyloplasts was conducted to gain further insights into the biochemical mechanisms involved in the grain-filling process [3]. A proteomic evaluation of the repercussions of heat stress on wheat grain-filling has also been conducted to determine its downstream effects on dough quality [4]. Proteomic approaches have also been utilized to determine the effects of chromosome deletion on protein expression in wheat seeds [5], and diploid, tetraploid, and hexaploid wheat flour proteomes have been evaluated to elucidate the effects of genome interaction on wheat proteins [6]. As noted, wheat proteomic research has been conducted on wheat seeds to ascertain target compounds and pathways for the improvement of value-added products, but not improvement of the agronomic properties of the plant itself.

Enhancement of wheat drought tolerance, durability to wind and cold, and pest-resistance will need to occur in the green leaf tissue and roots. Initial attempts were made to ascertain wheat leaf proteins induced by the aphid pest *Diuraphis noxia* (Mordvilko) in an attempt to find molecular markers associated with resistance [7], but to date neither leaf nor root tissue has been extensively characterized at the proteomic level. The focus of this research was to map and catalog the wheat leaf proteome to gain further insight into the biochemical makeup of wheat. This knowledge is the basic building block that will lead future researchers to potential target proteins whose addition or deletion could result in improvements to this and other agronomically important crops.

2 Materials and methods

2.1 Tissue

TXGBE307 hard red winter wheat was obtained from Dr. Mark Lazar at Texas A&M University. Wheat seed was planted individually in 3.8 cm diameter × 20.4 cm high Cone-tainers (Ray Leach Cone-tainer Nursery, Canby, OR, USA) containing Scotts Terra-Lite® Redi-earth® (Marysville, OH, USA). Cone-tainers were held in racks in water pans, with 48 seedlings *per* tray. Plants were grown in chambers with a 22:18°C day:night temperature cycle and a 14:10 day:night photoperiod until they reached the 1–2 leaf stage, approximately 7–10 d after planting. Each wheat seedling was harvested by cutting at the base of leaf number one, then quickly wrapping it in an aluminum foil pouch, and immediately submerging it in liquid nitrogen to minimize proteolytic activity. Wheat samples were stored at −80°C for no longer than 6 months. Seedlings from three racks were pooled and constituted one replication; three replications were analyzed.

2.2 Sample preparation

2.2.1 Water

Type I, 18 megaohom water further purified *via* glass distillation was used in the following experiments to alleviate yellowing and cloudiness during silver staining as well as to reduce contaminants that could potentially interfere with MALDI analysis and impede protein identification.

2.2.2 Protein precipitation

Wheat leaf tissue (5 g) was ground in liquid nitrogen with a ceramic mortar and pestle (Coors $2^3/_4$ in.). The resulting powder was suspended (1 g/5 mL) in chilled (−20°C) 10% TCA in acetone containing 0.07% β-mercaptoethanol (ME) and 1% plant protease inhibitor cocktail (P9599; Bio-Rad, Hercules, CA, USA),. The mixture was incubated at −20°C for at least 1 h then centrifuged at low speed (16 000 rpm) for 1 h. The pellet was washed three times (5 mL) with chilled (−20°C) acetone containing 0.07% ME and 1% plant protease inhibitor cocktail (Bio-Rad P9599) centrifuging at 16 000 rpm for 30 min between rinses. The fluid was removed and the pellet was dried slowly under nitrogen. If dried powder was not solubilized immediately, it was stored at −80°C for later use.

2.2.3 Protein solubilization

The wheat leaf proteins in the dried powder were solubilized in 8 M urea, 2% Triton X-100, and 60μmM DTT (30 mg:900 μL powder to solution, w/v) *via* incubation at 37°C for 1 h, vortexing every 15 min, ultrasonication with a microtip at 35% for 2 min (Sonic Dismembrator Model 300, Fisher, Hampton, NH, USA), followed by a final incubation at ambient temperature for 1 h. The mixture was centrifuged (45 000 rpm) (L8-M Ultracentrifuge; Beckmann, Coulter, Fullerton, CA, USA) for 1 h, and the supernatant harvested.

2.3 Protein quantification

A modified Bradford protein quantification assay was utilized to overcome interference of the 8 M urea and 60 mM DTT present in the solubilization solution [8]. Ovalbumin (Sigma, St. Louis, MO, USA) standard solutions (5–25 μg at 1 μg/μL) or 10 μL of sample was added to 10 μL 0.1 N HCl and 80 μL water. Bio-Rad's Protein Assay dye (500-0006) was

diluted with three volumes of water and mixed with the standards and samples (180 μL dye: 20 μL standard or sample) [9, 8]. Absorbance was read on a Bio-Rad Model 3550 Microplate Reader at 595 nm. The average protein concentration following solubilization (as previously described 30 mg powder: 900 μL resolubilization solution) was 3.6 μg/μL.

2.4 Analytical 2-DE

A four-gel system ((i) 4–7 pH IPG strip on an 11% SDS-polyacrylamide gel, (ii) a 6–11 pH strip on an 11% gel, (iii) a pH 4–7 strip on a 14% gel, and (iv) a pH 6–11 strip on a 14% gel) was utilized to enhance separation of wheat leaf proteins. IEF was performed using the Multiphor II (Amersham Biosciences, Piscataway, NJ, USA). Thirteen centimeter IPG strips (Amersham Biosciences), pH 4–7 and pH 6–11, were passively rehydrated overnight with 540 μg of protein in 250 μL of solubilization solution containing 2% carrier ampholyte (Pharmalyte pH 4–7 or 6–11; Amersham Biosciences). IEF of the acidic range IPG strips (pH 4–7) was conducted at 19°C for 3 h at 300 V and 18 h at 3500 V. IEF of the basic range IPG strips (pH 6–11) was conducted at 19°C for 3 h at 300 V and 21 h at 3500 V. Strips were subsequently stored at −80°C, or equilibrated and reduced in 50 mM Tris-HCl (pH 8.8), 6 M urea, 30% glycerol v/v, 2% SDS w/v, 65 mM DTT, and bromophenol blue for 15 min. Equilibrated strips were then placed on SDS-polyacrylamide gels, 16 cm × 20 cm, 11 or 14% acrylamide, and sealed with 0.5% agarose. SDS-PAGE was performed using the Protean II xi Cell, large gel format (Bio-Rad) at constant current (35 mA *per* gel) at 4°C until the bromophenol blue tracking dye was approximately 2–3 mm from the bottom of the gel. Three replicates were run for each gel to ascertain reproducibility.

2.5 Protein visualization

Proteins were visualized with silver stain using a modified version of Blum *et al.* [10]. Gels were fixed in 50% methanol and 12% acetic acid overnight, then rinsed with 50% ethanol (two times for 20 min) and water (20 min) before treating for 1 min with sodium thiosulfate (0.2 g/L). Gels were rinsed with water then incubated in silver nitrate (2.0 g/L) for 30 min. Incubated gels were rinsed with water and developed in a solution of sodium carbonate (60 g/L) and sodium thiosulfate (4.0 mg/L). Development was stopped with 5% acetic acid, and gels were stored in this solution until they could be processed and the reproducible spots removed from them. Three biological and three analytical replicates were analyzed. Protein spots were deemed reproducible if present in at least two of the biological as well as two of the analytical replicates. Protein number, p*I* and molecular weight were assigned using the PD-Quest gel analysis software (Bio-Rad).

2.6 In-gel digestion

Protein spots were removed from the gels and retained in 96-well microtiter plates. In-gel digestion of protein spots was conducted following a hybrid protocol developed from Jensen *et al.* [11], Shevchenko *et al.* [12], and the Keck Biotechnology Resource Laboratory at Yale University (www.info.med.yale.edu/wmkeck). A BioMek 2000 robot was utilized to perform the in-gel digestion, increasing throughput and reducing human error. For proteins of lower abundance, protein spots were removed from 3–4 2-D gels, pooled and digested in a single well of the 96-well microtiter plate. Protein spots were destained with 30 mM potassium ferricyanide and 100 mM sodium thiosulfate, then rinsed with 25 mM ammonium bicarbonate in 50% ACN according to the Yale protocol. The reduction and alkylation of the cysteine disulfide bonds were performed according to Jensen *et al.* [11] in 10 mM DTT and 55 mM iodoacetamide. The reswelling of the gel pieces and tryptic digestion of the proteins followed a slightly modified version of the Shevchenko *et al.* [12] protocol where the proteins were digested overnight at 37°C in 20 μL of 0.025 μg/μL trypsin (V5111; Promega, Madison, WI, USA) with no additional ammonium bicarbonate added. The supernatant was harvested the following day and the fluid further extracted from gel pieces with 0.1% TFA in 50% ACN (two times for 30 min) and then with 100% ACN (30 min). All extracted fluid was pooled with the trypsin supernatant and dried slowly under nitrogen to approximately 0.5–1.0 μL final volume.

2.7 MALDI-MS

Protein MS was conducted using a MALDI-TOF mass spectrometer (2000 Applied Biosystems DE-Pro, Applied Biosystems, Foster City, CA, USA). Digested protein samples were mixed (1:1 v/v) with a saturated solution of recrystallized CHCA (Sigma 14 550–5) matrix dissolved in 0.1% TFA/50% ACN then spotted on a MALDI plate (0.5 μL). A close external calibration was applied to all samples using Calibration Mixture 1 from Applied Biosystems, which entails a four-point calibration using bradykinin (904.4681), angiotensin I (1296.6853), Glu-fibrinopeptide B (1570.6774), and neurotensin (1672.9175). Samples were recalibrated using autolytic trypsin peaks (842.51 and 2211.10). The list of peptide masses from each PMF was saved for database analysis.

2.8 Bioinformatics

Monoisotopic peptide masses generated from the PMFs were used to search NCBI's wheat Unigenes using a local copy of ProteinProspector (version 3.2.1) and MS-Fit (version 3.1.1) running on a Windows NT4.0 server. The wheat Unigenes were downloaded from NCBI then formatted using the FA-Index program to create a protein molecular weight database

with indices for MS-Fit searches. The wheat Unigene set (Unigene Ta build 32) contained 22 306 entries with predicted peptide molecular weights ranging from 1000 to 94 000 Da with the longest protein of 824 amino acids. A total of 641 988 predicted trypsin fragments were generated with an average of 29 trypsin fragments *per* Unigene predicted peptide. If an acceptable match was not made to a sequence in the wheat Unigenes, the PMF peak list was used to search local versions of the rice, barley, and corn Unigenes, which were also downloaded monthly and formatted for MS-Fit. Acceptability criteria for searches are as follows: at least four peptide fragment masses matched the Unigene sequence and those peptide matches covered at least 10% of the putative protein sequence the Unigene coded for, and a BLASTX search conducted with the matched Unigene nucleic acid sequence must yield an E-value e-10 or less to be considered acceptable identification. If protein identification was not obtained from our local Unigene databases, the NCBI protein database was searched using ProFound (http://129.85.19.192/ profound_bin/WebProFound.exe) and MASCOT (http://www.matrixscience.com/cgi/index.pl?page=../home.html). Matches to protein sequences from the *Viridiplantae* taxon were considered acceptable if at least four peptide masses from the PMF matched, and a Z score of 1.00 or higher was obtained from ProFound or a significant score was obtained from MASCOT as *per* the program's algorithm, which rates scores as significant if they are above the 95% significance threshold ($p < 0.05$). Ideally, the molecular weight and p*I* of the protein identity obtained from the database agree with the experimentally obtained p*I* and molecular weight. However, some proteins successfully identified have substantial discrepancies between the experimental and database obtained p*I* and molecular weight, which can be caused by numerous factors such as PTMs, matches to broad protein class only, matches to proteins from different organisms, or genomic sequence, which could contain segments that are spliced out of the functional protein. Such protein identifications were deemed acceptable as long as the other statistical criteria were met.

3 Results and discussion

The four-gel system used to generate these reference maps lead to superior separation and focusing of the wheat leaf proteome, than is commonly achieved *via* single-gel analyses. Generation of these protein maps will enable future proteomic studies to focus on differential expression using these cataloged proteins as reference points, increasing throughput of later studies. In addition to providing a template for other wheat researchers, the results presented here show the increased feasibility of wheat leaf proteomics and perhaps plant proteomics in general. Plant protein and nucleic acid databases have grown substantially in the last few years yielding higher rates of successful identifications from mass spectrometric data. In total, 404 wheat leaf proteins were visualized using this 2-D PAGE four-gel system (Fig. 1). All reproducibly visualized proteins (277 in total) were assigned a number and were cataloged with their experimental p*I* and molecular weight. Proteins successfully identified or matched to genomic sequence are listed with their cataloging data in Table 1. Cataloging data, mass spectra and peak lists for those not identified may be viewed at http://entoplp.okstate.edu/labs/jwd/index.htm, and may be re-queried at a later date. Of the 277 cataloged proteins excised from the gels, 84 were identified querying the NCBI nonredundant protein database, and 58 were identified querying the local Unigene database sets followed by BLASTX search of the NCBI protein database for a 51% identification success rate.

The identified proteins were grouped by their functional annotation according to Bevan *et al.* [13] criteria (Fig. 2). Most of the proteins identified were involved in energy production/regulation and metabolism as would be expected in plant leaf tissue [14, 15]. The proteins involved in energy production (24%) include those which play a role in glycolysis, gluconeogenesis, the pentose phosphate pathway, the TCA cycle, respiration, fermentation, electron transport, and photosynthesis [13]. Proteins grouped under metabolism (1° = 12%; 2° =%) include those involved in the metabolism of amino acids, nitrogen and sulfur, nucleotides, phosphate, sugars and polysaccharides, lipids, sterols, and cofactors. Proteins classified as disease- and defense-related (12%) include resistance proteins, defense-regulated proteins, those involved with cell death, cell rescue, stress responses, and detoxification [13]. Storage proteins were also identified (4%), and some such as the glutelin precursor (169), a seed storage protein still present in these 7–10 d old seedlings, could be of interest to plant developmental biologists. These protein annotation ratios are comparable to the proteomic data published by Porubleva *et al.* [14] for maize, and by Watson *et al.* [15] for *Medicago truncatula*.

Database searching and bioinformatics have been reported to be one of the greatest stumbling blocks in proteomic research. Increases in the nucleotide sequence information available is due in large part to the completion of the *Arabidopsis* [16, 17] and rice [18–20] genomes, and as more plant genomic data is submitted and annotated the rate of successful protein identifications will increase. This is readily observable when examining the differences between data presented here and in the 2003 *M. truncatula* data [15] as compared to the 2001 maize proteomic study [14]. The percentage of protein matches to unannotated protein, gene, or EST sequences has dropped dramatically. In the 2001 maize study [14] 59% of the proteins were classified as unclear while the 2003 *M. truncatula* [15] study reported 3% and in this study only 8% were classified as unknown, hypothetical, or putative.

The maize [14] paper reported a 72% success rate, but of the 216 proteins identified, less than 50 were unique. Watson *et al.* [15] reported an identification success rate of 55%, utilizing both EST and protein databases, which is comparable

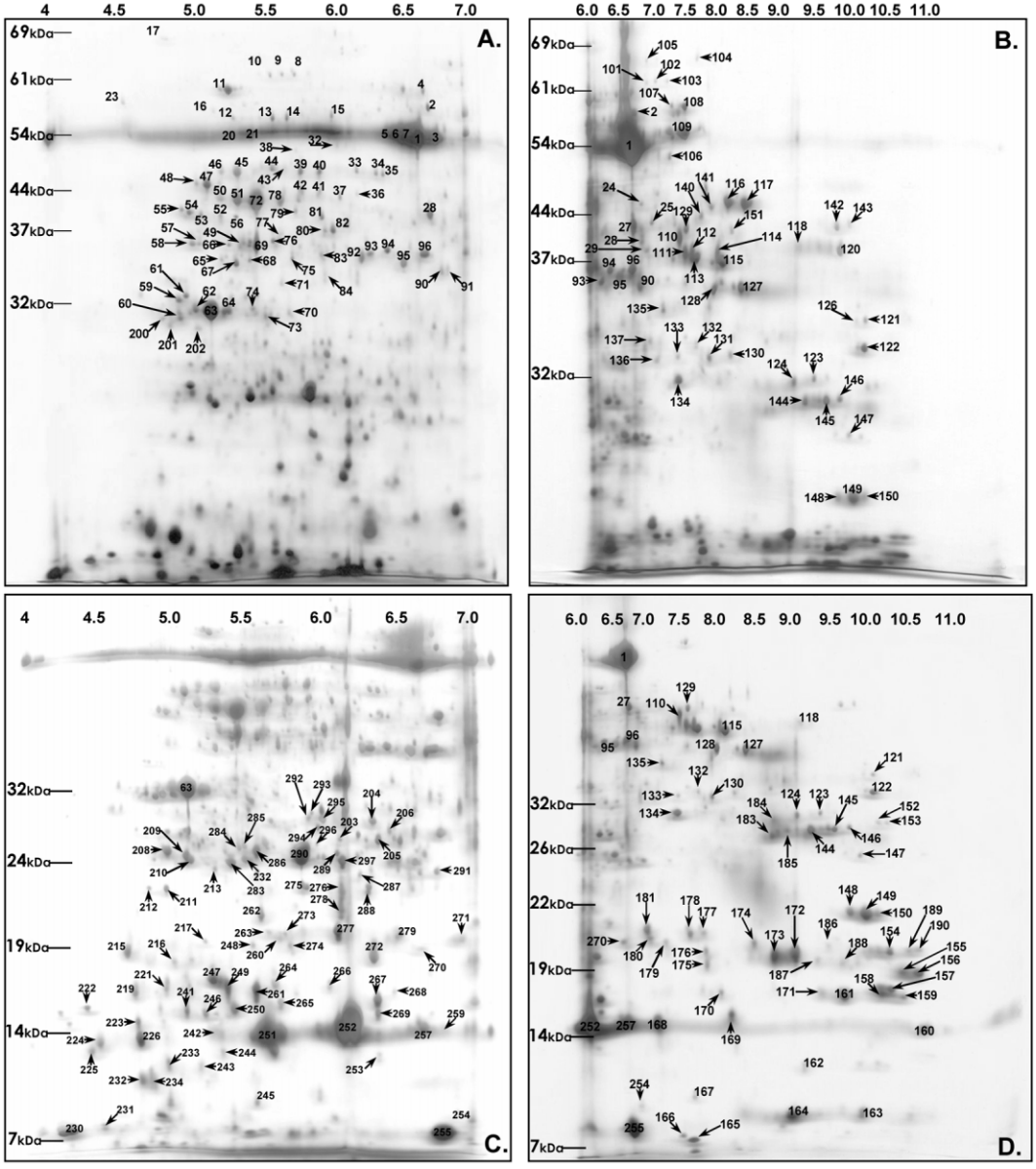

Figure 1. The four-gel system utilized to map the wheat leaf proteome. (A) 4–7 pH IPG strip, 11% SDS-PAGE, (B) 6–11 pH IPG strip, 11% SDS-PAGE, (C) 4–7 pH IPG strip, 14% SDS-PAGE, and (D) 6–11 pH IPG strip, 14% SDS-PAGE.

with the 51% identification success rate observed with the dual protein/EST search method utilized in this study. The *M. truncatula* study [15] exhibited slightly higher success due to their use of MS/MS data. Of the 142 proteins successfully identified in this study, 124 were unique. There are numerous reasons for multiple observations of the same protein on 2-D gels. The multiple spots could be isoforms with different signal or targeting sequences, which would cause shifts in p*I* and molecular weight. The proteins could be post-translationally modified where the addition of side chains, phosphate, methyl groups, *etc.* affect the p*I* and molecular weight. Protein degradation could also be responsible for multiple spots of the same protein, or as is the case with Rubisco, the protein could be carbamylated or merely overabundant and streaking. Many of these same phenomena are also responsible for the discrepancies observed between the experimentally determined and database observed p*I* and molecular weights.

As mentioned, the increase in successful identification/annotation is due in part to the amount of sequence and annotation data submitted to NCBI within the last few years. Since the beginning of 2001, the nucleotide database for flowering plants (*Magnoliophyta*) and *Viridiplantae* (all plants including algae) has increased seven-fold and the

Table 1. Proteins catalogued and identified from the wheat leaf proteome. The spot number from the 2-D SDS-PAGE, putative protein identification, the source organism the protein identity came from, the protein's accession number, the experimental p*I* and molecular weight as determined from the 2-D PAGE, the p*I* and molecular weight of the protein found in the database (calculated p*I* and mass), the database each identity was obtained from, the statistical score from the database, how many peptide fragments submitted to the database matched the protein identified (Pept Match), and the percentage of the protein's sequence those peptide fragments covered (% Cov) are listed below.

Spot no.	Putative protein ID	Protein source	Accession number	Experimental p*I*/mass (kDa)	Calculate p*I*/mass (kDa)	Database	Score	Pept match	% Cov
1	Rubisco large subunit	*Elyophorus globularis*	AAB82408	6.60/54.0	6.2/50.1	ProFound	2.43	9	15
2	Isoprene synthase	*Populus canescens*	CAC35696	6.70/58.5	5.3/68.9	ProFound	1.17	8	25
3	Rubisco large subunit	*Tacca palmata*	AAL37063	6.70/54.0	6.6/50.7	MASCOT	65	9	20
4	Rubisco large subunit	*Coleocarya gracilis*	AAD50092	6.60/61.0	6.4/51.6	MASCOT	93	11	22
5	Rubisco large subunit	*Kabuyea hostifolia*	CAA76746	6.30/54.0	6.5/50.0	ProFound	0.42	7	25
6	Rubisco large subunit	*Phragmites australis*	AAA61882.1	6.40/54.0	6.6/48.7	ProFound	0.12	4	7
7	Rubisco large subunit	*Isolepis bicolor*	CAC01208	6.50/54.0	6.3/52.5	ProFound	0.26	6	17
8	Fimbrin 1	*Arabidopsis thaliana*	AAB97846	5.65/65.0	6.1/67.8	ProFound	1.35	9	22
9	2,3-bisphosphogly-cerate-independent phosphoglycerate mutase	*Arabidopsis thaliana*	NP192690	5.55/65.0	5.5/61.0	EST/BLAST	e-131	8	27
10	2,3-bisphosphogly-cerate-independent phosphoglycerate mutase	*Arabidopsis thaliana*	NP192690	5.45/64.5	5.5/61.0	EST/BLAST	e-131	9	31
11	Reversibly glycosylated polypeptide	*Triticum aestivum*	CAA77237	5.20/61.0	5.8/41.5	EST/BLAST	0	4	16
12	H^+-transporting ATP synthase beta chain	*Triticum aestivum*	PWWTB	5.20/57.0	5.6/59.3	ProFound	2.32	13	32
13	Cytochrome P450	*Triticum aestivum*	AAR11387	5.50/57.0	5.5/77.4	EST/BLAST	0	5	16
14	At1g19370/F8O14_17	*Arabidopsis thaliana*	AAL67099	5.60/57.0	6.2/56.9	ProFound	1.10	8	24
15	Polyphenol oxidase (catechol oxidase)	*Ipomoea batatas*	CAA06855	5.95/57.0	5.8/55.3	ProFound	1.36	5	15
16	H^+-transporting ATP synthase beta chain	*Triticum aestivum*	PWWTB	5.05/58.0	5.1/53.8	MASCOT	102	20	44
17	Proliferating-cell nucleolar antigen	*Arabidopsis thaliana*	CAB80663	4.70/69.0	6.6/76.7	MASCOT	61	9	18
18	Phosphoenolpyruvate carboxykinase	*Oryza sativa*	AAP52715	4.80/61.0	6.3/71.4	ProFound	0.35	7	18
19	ATP synthase beta chain	*Aegilops columnaris*	BAA01870	5.15/58.0	5.2/53.9	ProFound	1.82	14	36
20	ATP synthase beta chain	*Aegilops columnaris*	BAA01870	5.05/54.0	5.2/53.9	ProFound	1.82	14	36
21	ATP synthase beta chain	*Triticum aestivum*	P20858	5.20/54.0	5.6/59.3	ProFound	2.43	10	47
23	Glucosyltransferase	*Nicotiana tabacum*	BAB60721	4.45/60.5	5.8/54.1	ProFound	2.43	8	22
32	BCS1 protein-like protein	*Arabidopsis thaliana*	AAM64718	6.00/53.5	6.1/55.0	ProFound	1.37	5	16
33	d-type cyclin	*Zea mays*	AAL83928	6.10/48.5	5.5/38.8	EST/BLAST	e-179	4	15
35	Mitogen-activated protein kinase	*Triticum aestivum*	AAO16559	6.30/48.5	5.7/70.6	EST/BLAST	0	7	16
36	Unknown protein	*Oryza sativa*	NP915536	6.10/45.0	6.3/70.6	EST/BLAST	e-151	4	13
38	Unknown protein	*Zea mays*	AAT42179	5.65/52.0	5.9/49.3	EST/BLAST	0	4	14
40	*S*-adenosylmethionine synthetase 2	*Hordeum vulgare*	BAA09895	5.80/48.5	5.5/42.8	EST/BLAST	0	6	26
41	DNA-binding protein 3	*Triticum aestivum*	NP188178	5.80/46.0	6.9/34.8	EST/BLAST	3 e-18	5	21
42	*S*-ribonuclease binding protein SBP1	*Arabidopsis thaliana*	AAG50626	5.65/46.0	5.2/37.5	ProFound	1.11	8	26
44	Eukaryotic translation initiation factor 4B	*Triticum aestivum*	AAC28254	5.50/49.0	5.7/47.6	EST/BLAST	e-178	5	13
46	Transcription factor-related	*Arabidopsis thaliana*	NP172466	5.15/48.5	4.4/44.9	MASCOT	47	8	28
47	G_2/mitotic-specific cyclin 2 (B-like cyclin)	*Oryza sativa*	Q40671	5.00/46.0	5.7/47.6	EST/BLAST	0	4	15
48	Protochlorophyllide reductase (ChlN subunit)	*Mesostigma viride*	NP038439	4.95/46.0	5.5/51.0	ProFound	0.36	9	32
49	Hypothetical protein	*Prunus armeniaca*	T51098	5.25/37.5	5.9/42.2	ProFound	1.56	7	24
50	Rubisco activase B	*Triticum aestivum*	AAF71272	5.10/43.5	6.9/47.8	EST/BLAST	0	5	13
51	Pathogen-related protein	*Triticum aestivum*	P16273	5.25/43.5	5.9/17.2	EST/BLAST	3 e-78	5	22
52	Sedoheptulose-1,7-bisphosphatase	*Triticum aestivum*	CAA46507	5.10/42.0	6/42.6	ProFound	1.31	10	29
54	Protoporphyrin IX magnesium chelatase	*Hordeum vulgare*	S64721	4.90/42.0	4.9/36.5	ProFound	1.36	9	35
58	Farnesyl pyrophosphate synthase	*Gossypium arboreum*	CAA72793	4.90/37.0	5.6/40.0	ProFound	0.64	7	22
59	Maturase K	*Mirabilis jalapa*	AAR20284	4.80/33.0	9.9/33.8	MASCOT	59	6	20
60	ATP synthase beta-subunit	*Pandorina morum*	BAB18833	4.90/31.5	5.5/40.8	ProFound	1.34	8	34

Table 1. Continued

Spot no.	Putative protein ID	Protein source	Accession number	Experimental pI/mass (kDa)	Calculate pI/mass (kDa)	Database	Score	Pept match	% Cov
61	Transducin family protein / WD-40 repeat family protein	*Arabidopsis thaliana*	NP176683	4.80/33.5	6.1/34.2	ProFound	0.26	6	22
62	Ras-related protein ARA-5	*Arabidopsis thaliana*	P28188	4.95/32.0	6.5/29.4	ProFound	0.73	6	51
63	Putative oxygen evolving protein of photosystem II	*Oryza sativa*	BAC21393	5.05/32.0	6.1/35.1	ProFound	1.38	6	23
66	Gibberellin 20-dioxygenase	*Triticum aestivum*	T06990	5.20/37.5	6.1/40.3	EST/BLAST	0	5	16
67	Putative plastidic cysteine synthase 1	*Oryza sativa*	NP914407	5.25/35.5	6.1/43.6	EST/BLAST	e-110	7	38
69	Caffeic acid O-methyltransferase	*Triticum aestivum*	AAQ07451	5.35/37.0	5.5/38.8	EST/BLAST	e-112	5	66
72	Rubisco activase	*Hordeum vulgare*	Q42450	5.30/43.5	5.6/47.5	ProFound	1.35	4	12
75	MYB40 – putative transcription factor	*Arabidopsis thaliana*	CAB87773	5.60/35.5	5.4/30.8	ProFound	0.51	5	28
77	ADP-glucose pyrophosphorylase	*Zea mays*	AAP47742	5.55/38.0	6.6/55.6	EST/BLAST	0	6	12
78	Ribulose-bisphosphate carboxylase activase	*Hordeum vulgare*	Q42450	5.55/43.5	5.6/47.5	EST/BLAST	0	6	16
80	26S proteasome regulatory particle triple-A ATPase	*Oryza sativa*	XP468146	5.90/38.0	8.9/47.2	EST/BLAST	0	7	23
82	Protein kinase family protein	*Arabidopsis thaliana*	NP196379	5.95/38.0	5.5/53.6	ProFound	1.32	8	19
83	*N*-acetylornithine deacetylase-like prot	*Arabidopsis thaliana*	CAB78785	5.85/36.0	5.1/44.5	ProFound	2.43	6	12
84	NADP-specific isocitrate dehydrogenase	*Oryza sativa*	NP917313	5.90/35.5	6.3/46.0	EST/BLAST	0	7	18
91	ATP synthase beta subunit	*Pinguicula lutea*	AAK72830	6.90/34.0	5.4/39.9	ProFound	2.43	11	38
92	Ribosomal protein L11	*Triticum aestivum*	BAB69029	6.15/36.0	9.3/37.6	EST/BLAST	e-109	6	26
94	Rubisco activase B	*Triticum aestivum*	AAF71272	6.35/ 36.5	6.9/47.8	EST/BLAST	0	5	16
95	GTP-binding protein	*Triticum aestivum*	AAP43929	6.55/35.5	8.4/68.0	EST/BLAST	e-110	5	14
96	Starch branching enzyme isoform RBE3	*Oryza sativa*	A48537	6.70/36.0	5.7/92.8	EST/BLAST	0	5	14
97	Glutathione *S*-transferase (GST6)	*Arabidopsis thaliana*	NP850479	10.00/21.0	8.5/29.3	ProFound	1.47	9	51
98	rps4	*Voitia hyperborea*	AAK83535	10.20/21.0	10.1/21.8	ProFound	2.43	4	23
102	H^+ transporting two-sector ATPase	*Triticum aestivum*	PWWTB	7.00/62.5	5.6/59.2	EST/BLAST	0	11	25
103	High-affinity phosphate transporter PT1	*Triticum aestivum*	AAD26146	7.20/62.0	8.8/43.5	EST/BLAST	0	4	10
109	Rubisco large subunit	*Triticum aestivum*	BAB47042	7.45/55.0	6.2/52.8	MASCOT	73	11	18
115	3-dehydroquinate dehydratase	*Oryza sativa*	NP918759	8.00/36.5	6.1/56.9	EST/BLAST	e-162	4	19
116	Ribulose bisphosphate carboxylase activase B	*Hordeum vulgare*	Q42450	8.10/44.5	7.6/47.2	EST/BLAST	0	4	12
117	Putative glucan synthase	*Oryza sativa*	BAB90325	8.35/44.5	8.8/190.8	EST/BLAST	4e-30	5	21
120	NADPH-protochlorophyllide oxidoreductase B	*Zea mays*	CAD99008	9.75/38.0	9.5/42.1	EST/BLAST	0	7	33
122	Arm repeat containing protein	*Triticum aestivum*	NP913815	10.10/30.0	8.3/28.8	EST/BLAST	9e-54	4	20
123	Outer mitochondrial membrane protein porin	*Triticum aestivum*	P46274	9.30/28.0	8.4/28.9	EST/BLAST	e-117	6	37
127	Malate dehydrogenase glyoxysomal precursor	*Oryza sativa*	Q42972	8.25/35.0	8.1/37.4	EST/BLAST	5e-77	4	27
129	Glyceraldehyde 3-phosphate dehydrogenase	*Hordeum vulgare*	P26517	7.50/39.5	6.7/36.1	ProFound	1.14	7	24
139	Putative 60S ribosomal protein	*Oryza sativa*	XP463021	9.75/24.0	10.6/25.2	ProFound	0.74	5	24
147	IB1C3–1 protein	*Arabidopsis thaliana*	CAA09808	9.85/24.5	9.6/28.3	ProFound	1.05	6	20
148	SERK1	*Helianthus annuus*	AAL93161	9.70/22.0	9.1/25.8	ProFound	0.95	7	41
149	Mitochondrial aldehyde dehydrogenase	*Oryza sativa*	BAA96793	9.95/22.0	6.3/58.9	EST/BLAST	0	5	14
150	Photosystem I reaction center subunit II	*Hordeum vulgare*	P36213	10.15/22.0	9.8/22.0	ProFound	1.71	7	33
155	Unknown protein	*Oryza sativa*	AF435650	10.50/18.0	6.9/55.8	EST/BLAST	0	5	14
160	Stripe rust resistance protein Yr1C	*Triticum aestivum*	AAG42168	10.80/14.0	7.2/93.2	EST/BLAST	0	7	14
164	NADPH-cytochrome P450 reductase	*Triticum aestivum*	AAG17471	9.00/8.0	5.0/73.0	EST/BLAST	0	6	16
168	Rubisco small subunit	*Triticum aestivum*	BAB19815	7.15/14.0	8.8/19.4	MASCOT	101	13	68
169	Glutelin precursor	*Oryza sativa*	BAD28254	8.20/14.5	9.2/56.3	EST/BLAST	0	4	20
172	Photosystem I chain IV Precursor	*Hordeum vulgare*	F1BH4	9.05/19.5	9.8/15.4	MASCOT	65	5	30
175	High-affinity phosphate transporter PT1	*Triticum aestivum*	AAD26146	7.75/18.5	9.0/60.1	EST/BLAST	0	6	18
181	Alternative oxidase	*Arabidopsis thaliana*	NP564395	7.00/21.0	6.3/33.1	ProFound	2.43	5	25

Table 1. Continued

Spot no.	Putative protein ID	Protein source	Accession number	Experimental pI/mass (kDa)	Calculate pI/mass (kDa)	Database	Score	Pept match	% Cov
182	Putative 60S ribosomal protein	*Oryza sativa*	XP463021	9.75/24.0	10.6/25.2	ProFound	0.74	5	24
200	Eukaryotic initiation factor 4A	*Oryza sativa*	BAA02152	4.70/31.5	5.5/47.1	EST/BLAST	0	5	19
201	Unknown protein	*Oryza sativa*	NP918035	4.75/31.0	4.7/48.0	EST/BLAST	e-117	5	29
202	Unknown protein	*Arabidopsis thaliana*	AAN38692	4.90/31.0	6.5/42.1	ProFound	2.43	5	19
203	Protein kinase	*Arabidopsis thaliana*	NP189510	6.30/26.0	9.6/42.0	MASCOT	55	10	55
205	Ascorbate peroxidase	*Hordeum vulgare*	CAA06996	6.45/26.0	5.8/27.5	ProFound	0.48	4	22
206	Alternative oxidase	*Triticum aestivum*	BAB88646	6.50/26.5	8.7/36.7	EST/BLAST	e-165	5	25
209	Putative calcium sensor protein	*Oryza sativa*	NP917878	5.10/24.5	5.0/31.4	ProFound	0.54	8	26
210	NADPH-cytochrome P450 reductase	*Triticum aestivum*	AAG17471	5.15/24.0	5.0/73.0	EST/BLAST	0	6	16
211	Cinnamyl-alcohol dehydrogenase	*Arabidopsis thaliana*	NP172422	4.90/22.5	6.8/39.4	ProFound	0.41	7	33
213	Hypothetical protein	*Oryza sativa*	BAB63631	5.30/23.5	10.6/52.7	MASCOT	59	8	30
214	RAS-related protein RAB2BV	*Beta vulgaris*	Q39434	4.70/21.5	6.4/23.9	ProFound	2.43	6	35
215	Cytochrome P450	*Triticum aestivum*	AAG17471	4.70/19.0	8.4/59.9	EST/BLAST	0	5	14
216	F-box protein family, AtFBX5	*Arabidopsis thaliana*	AAC31834	5.00/18.0	6.0/100.6	MASCOT	42	7	10
218	Hypothetical protein	*Arabidopsis thaliana*	NP174738	4.90/18.0	8.5/51.8	MASCOT	69	6	26
220	Calcineurin-like phosphoesterase-like protein	*Oryza sativa*	XP466494	4.05/17.0	4.5/20.0	EST/BLAST	5e-47	4	33
221	Expressed protein	*Triticum aestivum*	AAM13165	5.0/17.0	5.8/63.7	EST/BLAST	1e-33	5	23
224	Putative glycine decarboxylase subunit	*Triticum aestivum*	AAM92707	4.55/13.5	5.0/21.3	EST/BLAST	3e-83	5	46
226	RAS-related protein RAB7	*Glycine max*	Q43463	4.80/13.5	5.5/23.4	ProFound	2.43	5	35
228	Unknown protein	*Oryza sativa*	NP914887	5.20/16.5	5.4/19.5	EST/BLAST	3e-37	4	39
232	Putative phosphoenolpyruvate carboxykinase	*Oryza sativa*	AAM18765	4.80/11.0	6.3/71.2	ProFound	0.35	7	18
234	Thioredoxin family protein	*Arabidopsis thaliana*	NP973787	4.90/11.0	5.9/18.8	ProFound	0.96	5	39
241	Polyadenylate-binding protein	*Triticum aestivum*	T06979	5.10/15.5	4.8/19.0	ProFound	1.17	5	49
242	GTP-binding protein RAB11G	*Lotus japonicus*	CAA98183	5.40/14.0	5.2/24.6	ProFound	0.98	8	40
243	Protein import receptor TOM20, mitochondrial	*Solanum tuberosum*	T07679	5.25/12.0	5.3/22.8	MASCOT	59	7	56
245	V-ATPase G-subunit like protein	*Arabidopsis thaliana*	CAB79450	5.60/9.5	5.8/13.3	ProFound	2.21	9	57
247	Calmodulin	*Triticum aestivum*	P04464	5.30/17.5	4.1/16.1	ProFound	2.43	5	54
248	Pyruvate kinase-like protein	*Triticum aestivum*	T45821	5.55/19.5	6.3/53.4	EST/BLAST	2e-24	7	53
249	Origin recognition complex subunit 4	*Arabidopsis thaliana*	CAE01428	5.40/17.0	6.8/48.0	MASCOT	45	9	38
250	GTP-binding protein RAB1	*Glycine max*	S39565	5.45/15.5	5.3/22.7	ProFound	2.43	7	30
251	Rubisco small subunit	*Triticum aestivum*	BAB19815	5.70/14.0	5.8/13.3	ProFound	2.36	6	44
252	Rubisco small subunit	*Triticum aestivum*	BAB19815	6.20/14.0	5.8/13.3	ProFound	2.16	10	60
253	NBS-LRR-like protein	*Mentha longifolia*	AAL84890	6.40/13.0	5.3/20.3	ProFound	1.32	5	35
255	Putative RING zinc finger protein	*Arabidopsis thaliana*	AAM14996	6.85/7.5	7.0/12.7	ProFound	0.60	5	32
256	PRLI-interacting factor E	*Arabidopsis thaliana*	AAG31655	6.55/18.0	7.2/13.8	ProFound	1.22	5	68
257	Rubisco small subunit	*Triticum aestivum*	BAB19815	6.75/14.0	9.0/18.5	MASCOT	54	7	54
258	Small heat shock protein	*Triticum aestivum*	Q00445	6.85/15.5	6.2/23.5	EST/BLAST	1e-99	5	15
260	Glutathione *S*-transferase (GST Class-Zeta)	*Triticum aestivum*	O04437	5.65/19.5	6.1/24.0	ProFound	1.10	6	35
263	Unknown protein	*Triticum aestivum*	AAP03141	5.70/20.0	7.75/17.4	EST/BLAST	1e-47	11	38
264	Actin-depolymerizing factor 3	*Arabidopsis thaliana*	NP568915	5.75/17.0	5.7/15.9	EST/BLAST	3e-49	6	40
265	Triosephosphate-isomerase	*Hordeum vulgare*	AAB41052	5.80/16.0	5.4/27.0	ProFound	1.12	8	37
267	F-Box family protein	*Arabidopsis thaliana*	NP199913	6.45/16.0	5.2/16.3	ProFound	1.77	6	44
271	Ferritin 2 precursor	*Zea mays*	P29390	7.00/19.5	5.7/27.9	ProFound	1.41	5	26
272	MADS box transcription factor AP3–2	*Asarum europaeum*	AAF73927	6.40/18.5	5.8/24.5	ProFound	1.62	4	18
274	Protein kinase-like protein	*Oryza sativa*	XP468268	5.85/19.5	6.3/39.5	EST/BLAST	9e-95	4	25
276	Calcium-dependent protein kinase	*Oryza sativa*	XP483572	6.15/22.0	7.6/57.6	EST/BLAST	0	6	13
280	Rubredoxin putative	*Arabidopsis thaliana*	AAM63090	6.95/17.5	6.3/22.1	ProFound	1.19	4	29
283	Phosphatidylinositol bisphosphate phosphodiesterase	*Arabidopsis thaliana*	AAO63890	5.40/24.0	6.4/66.9	EST/BLAST	0	8	16
284	Proteasome subunit alpha type 2	*Oryza sativa*	AAT78811	5.45/25.5	5.4/25.8	EST/BLAST	e-128	6	23

Table 1. Continued

Spot no.	Putative protein ID	Protein source	Accession number	Experimental pI/mass (kDa)	Calculate pI/mass (kDa)	Database	Score	Pept match	% Cov
285	Triosephosphate-isomerase	*Hordeum vulgare*	P34937	5.50/25.5	5.4/27.0	ProFound	1.12	8	37
288	Beta-glucosidase	*Oryza sativa*	XP475121	6.35/22.5	6.9/58.5	EST/BLAST	0	6	16
289	NBS-LRR-like protein	*Oryza sativa*	NP919130	6.05/24.5	7.67/58.1	EST/BLAST	7e-47	5	10
291	Superoxide dismutase	*Triticum aestivum*	T06258	6.80/23.5	7.9/25.3	MASCOT	36	4	36
294	Putative selenocysteine methyltransferase	*Arabidopsis thaliana*	BAC42654	6.00/27.0	5.5/38.0	MASCOT	48	7	28
297	Dehydroascorbate reductase	*Triticum aestivum*	AAL71854	6.20/24.0	5.9/23.3	MASCOT	86	8	56

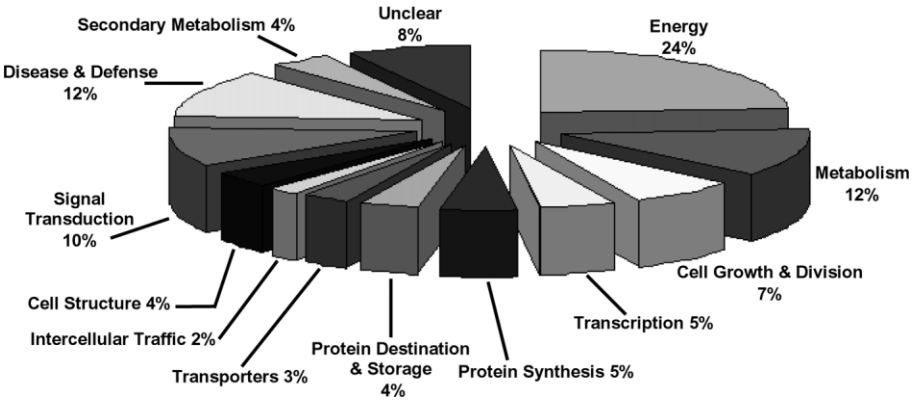

Figure 2. Functional annotation of wheat leaf proteome.

protein database almost five-fold. There are some taxonomic groups and species that are more heavily represented such as the grasses (*Poaceae*) compared to the legumes (*Fabaceae*). The *Poaceae* make up 54% of total plant nucleotide and 27% of the total protein sequence available in NCBI while the *Fabaceae* are only 10% of total nucleotide and 4% of the total protein sequence. The grasses are much more heavily represented, but the identification success rates were nearly identical. The similarity in successful identification rates in the *M. truncatula* study [15] and this research, and the fact that a large proportion of the identifications presented in the two are from different organisms, seems to indicate that plant protein sequence conservation is fairly high. These results indicate the increased feasibility of plant proteomics in general, and that proteomic techniques may be successfully applied to plant systems that are not well represented in the NCBI nucleic acid and protein databases.

4 Concluding remarks

The wheat proteome reference map presented here and at the supplemental website can be utilized for later comparative studies. In addition to the identified proteins with functional annotation, the nonidentified proteins provide a database/repository of cataloged information that may be resubmitted to the protein databases periodically as they continue to grow. The protein profiles will be useful for future comparisons to those generated during other wheat studies to ascertain at a glance those proteins affected by whatever perturbation of this system is being analyzed. The information could lead to the identification of biological markers for disease, insect resistance, and/or heat and drought tolerance. The reference maps give us the basic building block on which many crop improvement studies can be built. This study also showed that protein function ratios do not differ greatly between plant groups as vastly different as the moneocious grasses, corn and wheat, and the dioecious nitrogen-fixing legume *Medicago truncatula*. The protein identification success rates indicate the plant protein and nucleic acid databases are improving at an extremely rapid pace, and may eventually cease to be the limiting factor in the advancement of plant proteomics.

The data presented here can also be viewed in more detail at http://entoplp.okstate.edu/labs/jwd/index.htm. The four-gel system is presented with protein numbers hyperlinked to the cataloging data (p*I*, molecular weight and identification if obtained with all of the scoring data), as well as the PMF and mass peak list generated.

5 References

[1] Rosegrant, M., Ringler, C., Gerpacio, R., Paper presented at the XXIII International Conference of Agricultural Economists 1997, August 10–17, 1997, Sacramento, CA.

[2] Islam, N., Woo, S., Tsujimoto, H., Kawasaki, H. *et al.*, *Proteomics* 2002, *2*, 1146–1155.

[3] Andon, N., Hollingworth, S., Koller, A., Greenland, A. *et al.*, *Proteomics* 2002, *2*, 1156–1168.

[4] Majoul, T., Bancel, E., Triboi, E., Hamida, J. *et al.*, *Proteomics* 2003, *3*, 175–183.

[5] Islam, N., Tsujimoto, H., Hirano, H., *Proteomics* 2003, *3*, 307–316.

[6] Islam, N., Tsujimoto, H., Hirano, H., *Proteomics* 2003, *3*, 549–557.

[7] Porter, D. R., Webster, J. A., *Euphytica* 2000, *111*, 199–203.

[8] Ramagli, L., *Methods Mol. Biol.* 1999, *112*, 99–103.

[9] Ramagli, L. S., Rodriguez, L. V., *Electrophoresis* 1985, *6*, 559–563.

[10] Blum, H., Beier, H., Gross, H. J., *Electrophoresis* 1987, *8*, 93–99.

[11] Jensen, O. D., Wilm, M., Shevchenko, A., Mann, M., in: Link, A. J. (Ed.), *2-D Protoeme Analysis Protocols*, Humana Press, Totowa, New Jersey 1999, pp. 513–530.

[12] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.

[13] Bevan, M., Bancroft, I., Bent, E., Love, K. *et al.*, *Nature* 1998, *391*, 485–488.

[14] Porubleva, L., Vander Velden, K., Kothari, S., Oliver, D. J. *et al.*, *Electrophoresis* 2001, *22*, 1724–1738.

[15] Watson, B. S., Asirvathom, V. S., Wang, L., Sumner, L. W., *Plant. Physiol.* 2003, *131*, 1104–1123.

[16] The Arabidopsis Genome Initiative, *Nature* 2000, *408*, 816–820.

[17] Theologis, A., Ecker, J., Palm, C., Federspiel, N. *et al.*, *Nature* 2000, *408*, 816–820.

[18] Niiler, E., *Nat. Biotechnol.* 2000, *18*, 484.

[19] Goff, S., Ricke, D., Lan, T., Presting, G. *et al.*, *Science* 2002, *296*, 92–100.

[20] Yu, J., Hu, S., Wang, J., Wong, G. *et al.*, *Science* 2002, *296*, 79–92.

